International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Multimodal AI for Occupational Safety: A Quality-Aware Framework for Sensing, Risk Inference, Intervention, and Governance

Egberibo Paul Waredibamo*, Oluwaseun Omotoso**
*(Department of Human Physiology, Bayelsa Medical University, Yenagoa, Nigeria
Email: paulwaredibamoegberibo@gmail.com)
**(Department of Public Health, Lead City University, Ibadan, Nigeria
Email: oluwaseunomotoso95@gmail.com)

****

Abstract:
Artificial intelligence (AI) is increasingly used in occupational safety through computer vision, wearable sensing, Internet-of-Things (IoT) networks, equipment telemetry, and predictive analytics. Yet the literature remains fragmented across sensing modalities and frequently evaluates algorithmic accuracy without establishing whether AI deployment improves worker safety outcomes. This structured technical review synthesizes recent evidence from 2018–2026 and proposes a quality- and uncertainty-aware multimodal framework that connects sensing, edge processing, feature fusion, risk inference, selective decision-making, intervention, and governance. Recent reviews show rapid growth in PPE detection, worker-state monitoring, construction IoT, fatigue assessment, and accident prediction, but external validation, calibration, real-world robustness, and prospective outcome evaluation remain limited. To distinguish technological performance from safety effectiveness, a five-level evidence-maturity model is introduced: measurement validity (E1), algorithmic validity (E2), operational validity (E3), intervention effectiveness (E4), and worker health/safety outcomes (E5). A selective decision rule further distinguishes alert, no-alert, and indeterminate states when sensing quality or predictive confidence is insufficient. The framework links AI outputs to the hierarchy of controls and treats privacy, cybersecurity, worker autonomy, and human oversight as system-level requirements. The review concludes that AI can strengthen occupational risk detection and decision support, but high model performance should not be interpreted as evidence of reduced injury or illness without prospective workplace validation.

Keywords — artificial intelligence, occupational safety, computer vision, wearable sensors, Internet of Things, predictive analytics, uncertainty, hierarchy of controls

****

I. INTRODUCTION

Occupational injuries and work-related health risks remain persistent challenges across construction, manufacturing, transportation, energy, and other safety-critical industries. Conventional occupational health and safety (OHS) programs rely on hazard assessments, inspections, worker training, incident reporting, and established control measures to reduce exposure and prevent harm. Although these approaches remain fundamental, many workplace hazards are dynamic: worker–equipment interactions, environmental exposures, unsafe postures, physiological strain, and hazardous behaviors can develop rapidly and vary across workers, tasks, locations, and time. This has increased interest in digital technologies capable of providing continuous or near-real-time information to support occupational risk recognition and decision-making.

Artificial intelligence (AI), coupled with increasingly pervasive sensing technologies, provides new opportunities for such monitoring. Computer-vision systems can identify personal protective equipment (PPE), restricted-zone entry, unsafe worker–equipment proximity, postural hazards, and observable unsafe acts. Wearable and physiological sensors can characterize motion, location, fatigue-related responses, and physical strain, while connected environmental and equipment sensors can measure heat, gases, noise, particulate exposure, machine state, and other contextual variables. Machine-learning models can subsequently integrate these heterogeneous data streams to detect hazards, recognize unsafe states, and estimate emerging

risk. Recent reviews document rapid growth in AI applications across construction safety, wearable sensing, Internet-of-Things (IoT) monitoring, accident analytics, and broader AI-enabled OHS [1]–[7].

Despite this technological progress, a fundamental gap remains between **algorithmic performance and demonstrated occupational-safety effectiveness**. Most published systems are evaluated using intermediate technical endpoints such as classification accuracy, F1-score, area under the receiver-operating-characteristic curve, or mean average precision. These metrics establish whether an algorithm can recognize a predefined target under particular experimental conditions, but they do not demonstrate that deployment reduces hazardous exposure, injuries, or occupational illness. This distinction is evident in the systematic review by Jetha *et al.* [8], which identified 1,255 records concerning AI-enabled OHS tools published between 2018 and 2024 but found only two studies meeting its criteria for quantitative worker morbidity or mortality outcomes; neither was assessed as high quality. The literature therefore reveals a substantial **translation gap from sensing and prediction to intervention and measurable worker outcomes**.

A related challenge is the reliability of AI systems outside controlled development environments. Computer-vision performance can deteriorate because of occlusion, illumination changes, camera geometry, unfamiliar equipment, PPE variation, and domain shift [3], [9]. Wearable systems introduce additional uncertainty associated with sensor placement, motion artifacts, calibration, battery limitations, user compliance, and signal quality [4]–[6]. IoT-based monitoring further depends on wireless coverage, sensor synchronization, interoperability, network availability, and cybersecurity [7]. Consequently, the absence of a detected hazard cannot always be interpreted as evidence of a safe state. In safety-critical applications, a system should distinguish between **"no hazard detected"** and **"hazard status cannot be determined reliably."**

These limitations also expose a broader systems-level problem. Occupational hazards are rarely represented completely by a single sensing modality. A camera may observe worker–equipment proximity but not physiological strain; a wearable may detect abnormal motion but not an environmental hazard; and an environmental sensor may identify excessive exposure without revealing which workers are affected. Multimodal sensing can provide complementary information, but simply combining additional sensors does not guarantee reliable inference. The quality, availability, uncertainty, and temporal consistency of each modality must also be considered. A deployment-oriented AI architecture should therefore integrate **multimodal sensing, explicit data-quality assessment, uncertainty-aware inference, and selective decision-making**, allowing the system to abstain or escalate for human review when available evidence is insufficient.

This systems perspective is consistent with reliability principles used in other safety-related engineering domains. Deep-learning-based signal classification for pipeline structural-health monitoring [10] and physics-aware machine learning for leak detection and quantification [11] illustrate how sensor measurements, signal processing, model structure, and validation can be integrated to infer system condition from imperfect observations. Similarly, engineered frequency selectivity in UWB sensing hardware [12] illustrates the broader importance of suppressing interfering information before downstream interpretation. These studies do not constitute occupational-health evidence; rather, they provide methodological analogies for designing sensing and inference systems in which measurement quality and model reliability are consequential.

The remaining challenge is therefore not simply to develop more accurate AI classifiers, but to determine **how AI-generated information progresses from measurement to prediction, from prediction to intervention, and ultimately from intervention to demonstrable safety benefit**. Existing literature is distributed across computer vision, wearable sensing, IoT, physiological monitoring, predictive analytics, and accident intelligence, while validation is frequently concentrated at the sensor or algorithm level. A unified framework is needed to distinguish these evidence levels, identify where uncertainty enters the decision chain, and connect AI outputs to established occupational-risk-management principles, including the hierarchy of controls.

Accordingly, this review makes three principal contributions. First, it synthesizes recent evidence

on AI-enabled occupational safety across computer vision, wearable and physiological sensing, IoT and environmental monitoring, and predictive risk analytics. Second, it introduces a **five-level evidence-maturity framework** that distinguishes measurement validity (E_1), algorithmic validity (E_2), operational validity (E_3), intervention effectiveness (E_4), and worker health and safety outcomes (E_5), thereby preventing algorithmic performance from being conflated with demonstrated safety effectiveness. Third, it proposes a **quality- and uncertainty-aware multimodal safety architecture** that integrates sensing quality, risk estimation, predictive confidence, selective decision-making, hierarchy-of-controls intervention, human oversight, and governance. Together, these contributions provide a structured pathway for moving AI-enabled occupational safety research from isolated detection and prediction tasks toward reliable, responsible, and outcome-oriented workplace systems.

II. REVIEW METHODOLOGY

A. Review Design and Scope

This study employed a structured technical review to synthesize recent developments in artificial intelligence (AI)-enabled occupational safety and to identify methodological and translational gaps across sensing, inference, intervention, and deployment. The review focused primarily on literature published between **2018 and 2026**, a period characterized by rapid adoption of deep learning, computer vision, wearable sensing, Internet-of-Things (IoT) platforms, and multimodal analytics in occupational health and safety (OHS). Because the objective was to integrate heterogeneous technological and safety evidence rather than estimate a common intervention effect, the study was designed as a structured evidence synthesis rather than a statistical meta-analysis.

B. Search Strategy and Source Selection

The literature search targeted peer-reviewed journal articles, systematic and scoping reviews, and representative applied studies addressing AI or intelligent sensing for occupational safety. Authoritative guidance from organizations such as the National Institute for Occupational Safety and Health (NIOSH), International Labour Organization (ILO), and European Agency for Safety and Health at Work (EU-OSHA) was additionally considered for established safety-control and governance principles.

Search terms were organized around three concept groups. The first described the application context, including *occupational safety*, *workplace safety*, *construction safety*, *worker monitoring*, *occupational health*, and *hazard detection*. The second captured enabling technologies, including *artificial intelligence*, *machine learning*, *deep learning*, *computer vision*, *wearable sensors*, *physiological sensing*, *Internet of Things*, *natural-language processing*, and *multimodal sensing*. The third addressed safety and deployment outcomes, including *accident prediction*, *fatigue*, *risk assessment*, *near miss*, *uncertainty*, *sensor quality*, *privacy*, *human oversight*, and *worker safety outcomes*. Terms were combined iteratively to identify both technology-specific studies and literature examining broader OHS implications.

C. Eligibility and Evidence Extraction

Studies were prioritized when they addressed an identifiable occupational or industrial safety problem and provided sufficient information regarding at least one of the following: sensing modality, AI or analytical method, prediction or detection task, validation strategy, deployment environment, operational limitation, intervention, or worker-safety outcome. Studies concerned exclusively with general-purpose AI without an occupational-safety application were excluded from the principal evidence synthesis. Engineering studies outside OHS were retained only where explicitly identified as methodological analogies and were not treated as evidence of occupational-safety effectiveness.

For each relevant study, evidence was extracted according to the **industrial sector, safety objective, sensing/data modality, analytical method, validation setting, reported performance criteria, deployment characteristics, and principal limitations**. Particular attention was given to whether evaluation was performed on laboratory or workplace data, whether workers, sites, equipment, or time periods were independently held out during validation, and whether the study evaluated technical performance alone or reported an intervention or worker-level safety outcome.

D. Evidence Synthesis and Maturity Classification

The extracted literature was synthesized along three complementary dimensions. First, studies were categorized by **safety function** as detection, prediction, decision support, or intervention. Second, they were classified by **information modality**, including computer vision, wearable/inertial sensing, physiological sensing, environmental monitoring, localization, equipment telemetry, textual data, and multimodal combinations. Third, the strength of evidence was assessed according to the stage of the safety decision chain addressed by each study.

Because the reviewed literature encompasses heterogeneous hazards, sensing technologies, datasets, prediction targets, and performance metrics, quantitative effect sizes were not pooled. Instead, evidence was interpreted using the five-level maturity framework developed in Section V: **measurement validity (**E1E_1**), algorithmic validity (**E2E_2**), operational validity (**E3E_3**), intervention effectiveness (**E4E_4**), and worker health and safety outcomes (**E5E_5**)**. This classification enables technical model performance to be distinguished explicitly from evidence that an AI-enabled system operates reliably in the workplace, changes safety-relevant processes or exposures, or ultimately improves worker outcomes.

III. AI-ENABLED OCCUPATIONAL SAFETY LANDSCAPE

A. Computer Vision and Spatial Hazard Detection

Computer vision has emerged as one of the most extensively investigated AI technologies for occupational-safety monitoring because many workplace hazards have directly observable spatial or behavioral signatures. Deep-learning-based systems have been applied to personal protective equipment (PPE) compliance, restricted-zone intrusion, worker–equipment proximity, posture and activity recognition, vehicle and machinery detection, and identification of selected unsafe acts. These capabilities are particularly relevant in dynamic environments such as construction sites, manufacturing facilities, warehouses, and other workplaces in which manual observation cannot provide continuous spatial coverage. Recent systematic evidence demonstrates substantial growth in deep-learning-based PPE and hazard-recognition systems [3], while earlier critical reviews show that construction-safety applications have predominantly concentrated on object detection, tracking, and activity recognition [9].

High recognition performance under benchmark conditions, however, does not necessarily translate into reliable hazard detection during deployment. Occlusion, variable illumination, dust, weather, camera orientation, worker density, changing site geometry, and previously unseen PPE or equipment can alter the visual distribution encountered by a trained model. More importantly, conventional metrics such as mean average precision (mAP) characterize object-detection performance but do not fully describe operational safety performance. Deployment-oriented evaluation should therefore consider event-level sensitivity, false alarms per worker-hour or camera-hour, missed-event severity, detection latency, predictive calibration, and generalization across workers, sites, and operating conditions. Computer vision should consequently be regarded as a powerful but context-dependent source of safety information rather than an independently sufficient representation of workplace risk.

B. Wearable and Physiological Sensing

Wearable technologies complement vision by providing worker-centered information that may remain unavailable to fixed or remote cameras. Inertial measurement units (IMUs), physiological sensors, smart garments, localization devices, and wearable environmental monitors have been investigated for activity recognition, ergonomic assessment, proximity monitoring, fatigue-related responses, physiological strain, and exposure characterization [4]–[6]. Wang *et al.* [5], in a review of 200 publications, identified sensor selection and placement, experimental validity, multimodal data fusion, end-to-end analytics, worker-state modeling, and human–technology interaction as continuing research challenges. Physiological sensing has similarly been investigated for fatigue, stress-related responses, attention, and other worker states relevant to safety [4].

Wearable monitoring nevertheless introduces a different set of reliability constraints. Measurements can be affected by sensor

displacement, motion artifacts, poor skin or garment contact, calibration error, battery depletion, communication loss, environmental conditions, and worker compliance. Applied studies are increasingly evaluating wearable measurements under more realistic work conditions, including machine-learning analysis of IMU signals for construction-worker fatigue trends [13]; however, much of the available evidence remains laboratory-based, task-specific, or derived from relatively constrained populations. Generalization to unseen workers and work environments therefore requires explicit evaluation. Physiological and behavioral signals should also be interpreted as indicators of safety-relevant state rather than direct clinical diagnoses unless independently validated for that purpose.

An important implication is that **data availability and data validity are distinct conditions**. A wearable device may continue producing numerical values even when the underlying signal is corrupted. Conversely, missing data should not be interpreted as evidence of normal worker status. Signal-quality assessment must therefore accompany downstream inference in safety-critical wearable systems.

C. IoT, Environmental Sensing, and Equipment Telemetry

Internet-of-Things architectures extend occupational monitoring from individual sensors to distributed cyber-physical systems in which worker, equipment, environmental, and site information can be acquired and communicated in near real time. A review encompassing 147 publications categorized construction-safety IoT applications primarily into worker/labor monitoring, environmental monitoring, and site or activity monitoring [7]. Typical measurements include temperature, humidity, gases, particulate matter, noise, equipment state, worker location, proximity, and other contextual variables that may precede hazardous events.

IoT sensing is particularly valuable when used in conjunction with other modalities. Computer vision provides rich spatial information but is constrained by camera coverage and visibility, whereas connected sensors can characterize environmental or equipment conditions outside the visual field [14]. Equipment telemetry can further reveal machine states that may be difficult to infer reliably from images alone. These complementary capabilities create a stronger basis for contextual risk assessment than isolated sensing.

The resulting architecture, however, introduces additional system-level failure modes. Sensor synchronization, wireless coverage, communication latency, battery life, device ruggedization, interoperability, calibration, maintenance, and cybersecurity can all influence the reliability of the final safety decision. Consequently, validation should extend beyond individual sensor or classifier accuracy to the availability, timing, integrity, and resilience of the complete sensing-to-decision pipeline.

D. Predictive Analytics and Accident Intelligence

Whereas vision, wearable, and IoT systems primarily characterize current workplace conditions, predictive analytics attempts to estimate **future safety risk** from historical and time-varying information. Machine-learning models have been developed using incident records, near-miss reports, task characteristics, worker-related variables, environmental conditions, equipment states, and organizational data to identify patterns associated with occupational injury and other adverse events [15]. Depending on the application, predictive models may support worker-level risk estimation, task prioritization, inspection planning, or identification of combinations of conditions associated with elevated incident probability.

Unstructured safety records provide an additional source of information. Natural-language processing (NLP) and pretrained language models can transform accident narratives, investigation reports, and near-miss descriptions into structured representations for accident classification, causal-theme extraction, and risk-pattern discovery [16]. Such approaches can help organizations learn from large repositories of historical reports that would otherwise require labor-intensive manual analysis.

Predictive performance must nevertheless be interpreted cautiously. Occupational incidents—particularly severe events—are comparatively rare, producing substantial class imbalance and potentially misleading accuracy estimates. Models trained on historical data may also reproduce reporting practices, organizational biases, or site-specific relationships rather than transferable causal mechanisms. Evaluation should therefore

emphasize precision–recall behavior, sensitivity to severe events, calibration, temporal and site-based validation, and decision-relevant false-alarm costs. Most importantly, accurate retrospective prediction establishes **predictive validity**, not evidence that deploying the model prevents an incident. Demonstrating safety effectiveness requires an additional intervention and outcome-evaluation stage.

E. Cross-Modality Synthesis

The reviewed technologies provide complementary rather than interchangeable information. Vision is well suited to spatial and observable behavioral hazards; wearable systems provide worker-centered motion and physiological information; environmental and IoT sensors characterize exposure and contextual conditions; equipment telemetry represents machine state; and predictive or textual analytics provide temporal and historical risk information. Their limitations are similarly modality specific.

This complementarity motivates multimodal occupational-safety systems, but multimodality alone does not guarantee reliability. Combining a high-quality modality with corrupted, unavailable, or domain-shifted measurements can degrade rather than improve inference. The central technical problem therefore shifts from simply **acquiring more data** to determining **which information is trustworthy, how complementary evidence should be fused, and when the resulting prediction is sufficiently reliable to support intervention**. These requirements motivate the quality- and uncertainty-aware framework developed in Section IV.

AI Domain	Typical Safety Applications	Evidence Status	Key Gap
Computer vision [3], [9]	PPE, proximity, posture, hazard detection	Extensive detection studies	Domain shift and field validation
Wearable sensing [4]–[6], [13]	Fatigue, ergonomics, activity, worker state	Growing laboratory and field evidence	Signal quality and worker generalization
IoT and telemetry [7], [14]	Exposure, environment, equipment and site monitoring	Increasing integrated applications	Connectivity and system reliability
Predictive/NLP analytics [15], [16]	Injury prediction, accident classification, risk mining	Strong retrospective applications	Prospective validation and class imbalance
Safety outcomes [8]	Injury and illness prevention	Limited direct outcome evidence	Intervention and worker-outcome validation

IV. QUALITY- AND UNCERTAINTY-AWARE MULTIMODAL FRAMEWORK

A. Multimodal Safety-State Representation

Occupational risk is inherently multimodal. A hazardous state may depend simultaneously on worker behavior, physiological condition, location, environmental exposure, and equipment operation. Computer vision, wearable sensors, environmental monitors, localization systems, and equipment telemetry therefore provide complementary rather than interchangeable information.

For worker i at time t, let $\mathbf{x}_{i,t}$ represent the available multimodal observations and $\mathbf{q}_{i,t}$ their corresponding data-quality information. The proposed framework estimates future risk over prediction horizon Δ using

$$\hat{\mathbf{r}}_i(t, \Delta) = f\theta(\mathbf{x}_{i,t-L:t}, \mathbf{q}_{i,t-L:t}) \quad (1)$$

where L is the observation window and $f\theta$ is the risk-estimation model. The observation vector may contain visual, motion, location, physiological, environmental, and equipment/site information, while the quality vector represents factors such as sensor availability, missing data, image visibility, signal integrity, synchronization, and sensor health.

The central principle is that **sensor availability is not equivalent to sensor reliability**. A camera may be occluded, a wearable signal may contain motion artifacts, or an environmental sensor may temporarily lose communication while still producing incomplete information. Such conditions should influence the confidence assigned to the resulting prediction rather than being interpreted as evidence of low risk.

Multimodal fusion may occur at the feature, representation, or decision level. Regardless of the implementation, the contribution of each modality should reflect both its predictive value and its quality at the time of inference.

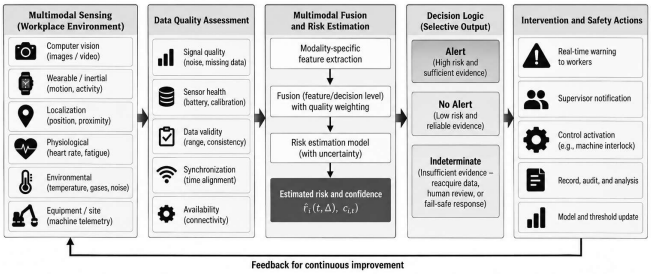

FIG. 1. QUALITY- AND UNCERTAINTY-AWARE MULTIMODAL ARCHITECTURE FOR AI-ENABLED OCCUPATIONAL SAFETY.

B. Quality, Uncertainty, and Selective Decisions

Data quality and predictive uncertainty should be treated as related but distinct quantities. **Data quality** concerns the reliability of the measurements supplied to the model, whereas **predictive uncertainty** concerns confidence in the resulting inference. A high-quality observation can still yield an uncertain prediction under domain shift or an unfamiliar workplace condition, while degraded measurements can increase uncertainty even within a familiar operating environment.

For this reason, the proposed framework does not force every observation into a binary safe/unsafe decision. Instead, three operational states are used:

Alert — estimated risk exceeds the intervention threshold and the available evidence is sufficiently reliable.

No Alert — estimated risk remains below the intervention threshold and both data quality and predictive confidence are adequate.

Indeterminate — data quality or predictive confidence is insufficient to support either conclusion reliably.

The third state is particularly important in safety-critical applications because **"no hazard detected" is not equivalent to "hazard status cannot be determined."** An indeterminate result may trigger data reacquisition, another sensing modality, human review, or a predefined fail-safe response. Deployment evaluation should therefore consider calibration, false-alarm burden, missed-event severity, latency, data quality, and uncertainty in addition to conventional accuracy metrics.

V. FROM ALGORITHM PERFORMANCE TO SAFETY OUTCOMES

A recurring limitation in AI-enabled occupational safety research is the use of algorithmic performance as a proxy for safety effectiveness. Metrics such as F1-score, mAP, and AUROC can demonstrate detection or predictive capability under specified test conditions, but they do not establish that deployment reduces hazardous exposure, injury, or illness. To distinguish these claims, Fig. 2 organizes evidence into five maturity levels, progressing from **measurement validity (E1)** and **algorithmic validity (E2)** to **operational validity (E3)**, **intervention effectiveness (E4)**, and ultimately **worker health and safety outcomes (E5)**.

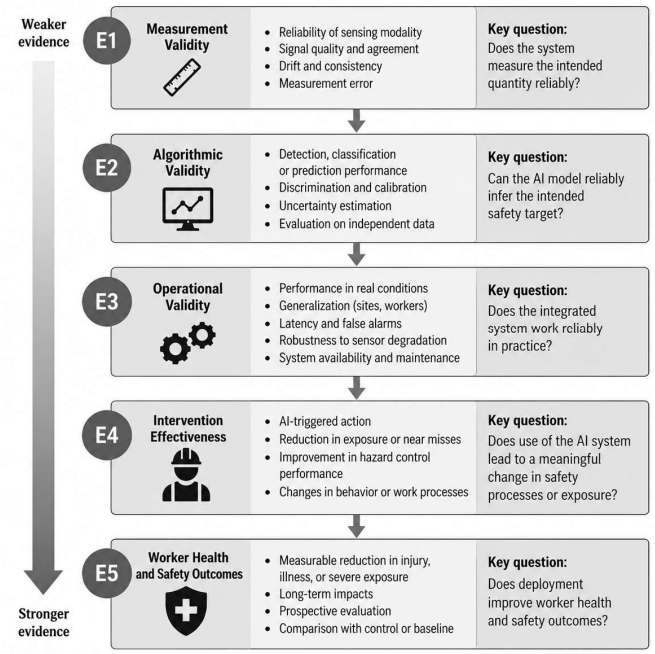

FIG. 2. FIVE-LEVEL EVIDENCE-MATURITY MODEL FOR AI-ENABLED OCCUPATIONAL SAFETY.

The levels represent progressively stronger evidence rather than interchangeable performance criteria. Success at E2, for example, establishes predictive capability but does not by itself demonstrate operational reliability (E3), effective intervention (E4), or improved worker outcomes (E5). This distinction is particularly important given the review by Jetha *et al.* [8], which identified only two eligible studies directly evaluating worker morbidity or mortality outcomes. Future evaluations should therefore extend beyond model accuracy toward real-world validation, intervention effectiveness, and prospective worker-outcome assessment.

TABLE II
EVIDENCE-MATURITY FRAMEWORK

Level	Evidence Focus	Typical Evaluation
E_1	Measurement validity	Signal quality, agreement, drift
E_2	Algorithmic validity	F1, AUPRC, mAP, calibration
E_3	Operational validity	Generalization, latency, false alarms
E_4	Intervention effectiveness	Exposure or hazard reduction
E_5	Worker outcomes	Injury, illness, severe exposure

The levels represent progressively stronger evidence, but performance at one level does not establish the next. In particular, strong algorithmic performance at E_2 cannot be interpreted as

evidence of improved worker outcomes at E_5. This distinction is supported by Jetha *et al.* [8], whose systematic review found only two eligible studies directly evaluating worker morbidity or mortality outcomes despite a much larger literature on AI-enabled OHS tools. Stronger research should therefore progress from technical validation toward real-world deployment, intervention assessment, and prospective worker-outcome evaluation.

VI. INTEGRATION WITH THE HIERARCHY OF CONTROLS

AI-enabled monitoring should complement established occupational-risk management rather than become an additional layer of worker surveillance. The NIOSH hierarchy of controls prioritizes elimination, substitution, and engineering controls over administrative controls and PPE because higher-order controls reduce reliance on individual behavior [17]. Accordingly, AI should be used not only to detect unsafe conditions, but also to identify recurring risk patterns that can inform more effective hazard control.

Repeated alerts can provide evidence of underlying deficiencies in the work system. Persistent worker–vehicle proximity events, for example, may indicate the need for route separation, physical barriers, traffic redesign, or automated interlocks rather than additional warnings. Similarly, recurring PPE non-compliance may reflect problems with equipment suitability, fit, thermal burden, workflow, or an underlying hazard that could be addressed through higher-order controls.

AI should therefore support a closed-loop process in which detected risks inform root-cause assessment, appropriate control selection, and subsequent evaluation of whether exposure has been reduced. This positions AI as a decision-support component within the hierarchy of controls rather than a substitute for safe work design. Feedback from implemented controls can subsequently refine sensing strategies, risk thresholds, and intervention policies.

VIII. FAILURE MODES AND VALIDATION REQUIREMENTS

AI-enabled safety systems can fail at multiple points between sensing and intervention. Sensor degradation, missing modalities, domain shift, poor calibration, communication delays, and excessive false alarms can undermine otherwise accurate models. Validation should therefore address the complete sensing-to-decision pipeline rather than algorithmic accuracy alone. Table III summarizes the principal failure modes and corresponding safeguards.

TABLE III
FAILURE MODES AND REQUIRED SAFEGUARDS

Failure Mode	Potential Effect	Required Safeguard
Missing or degraded sensor data	Unreliable risk estimate	Quality checks and fallback sensing
Sensor drift or miscalibration	Systematic measurement error	Calibration and sensor-health monitoring
Domain shift	Reduced field performance	External and cross-site validation
Poor model calibration	Misleading confidence	Calibration and uncertainty assessment
Excessive false alarms	Alarm fatigue and reduced trust	Threshold tuning and event-level evaluation
Communication or processing delay	Late safety response	Latency and availability testing
Data leakage during validation	Inflated performance estimates	Worker-, site-, or time-based holdout
Ambiguous model output	Unsafe automated decision	Human review or fail-safe escalation

Validation should be conducted at progressively broader levels. **Sensor-level validation** should assess signal quality, missingness, drift, calibration, and synchronization; **model-level validation** should examine discrimination, calibration, robustness, and uncertainty; and **system-level validation** should evaluate latency, availability, false-alarm burden, and escalation behavior. Finally, **deployment-level evaluation** should determine whether the integrated system performs reliably under realistic workplace conditions and produces measurable improvements in exposure, near misses, or worker outcomes.

Particular attention is required when constructing training and test sets. Randomly distributing frames or overlapping time windows from the same worker, site, shift, or recording session across both sets can introduce information leakage and produce overly optimistic performance estimates. Validation should instead hold out the unit expected to be novel during deployment, such as an unseen worker, site, project, equipment configuration, or time period.

IX. RESEARCH GAPS AND FUTURE DIRECTIONS

Despite rapid progress in AI-enabled occupational safety, several challenges continue to limit reliable workplace deployment. Multimodal learning requires more robust strategies for handling missing, degraded, and asynchronous sensor streams rather than assuming that all modalities remain continuously available. Rare but high-consequence events also require evaluation beyond conventional accuracy measures, with greater emphasis on precision–recall performance, event-level detection, calibration, and the asymmetric consequences of missed hazards and false alarms. Uncertainty-aware inference should consequently become an integral component of safety-critical AI rather than an optional model feature.

Greater attention is also needed to external and prospective validation. Models should be evaluated across unseen workers, sites, equipment, and operating conditions to establish robustness to domain shift. Privacy-preserving edge and distributed approaches offer opportunities to reduce unnecessary transfer of identifiable worker data, but their benefits should be evaluated alongside system reliability, computational constraints, cybersecurity, and governance requirements. Most importantly, future studies should progress beyond algorithmic performance to determine whether AI-supported interventions measurably reduce hazardous exposure, near misses, injuries, or occupational illness.

Future research should also examine the organizational consequences of AI monitoring. A technically accurate system may provide limited safety benefit if it reduces worker trust, discourages incident reporting, produces systematically unequal false alarms, or shifts responsibility for inadequately controlled hazards toward individual workers. Technical performance should therefore be evaluated alongside worker participation, transparency, perceived fairness, human oversight, and acceptance. Addressing these technical and organizational dimensions together will be essential for moving AI-enabled occupational safety from experimental monitoring systems toward reliable and responsible workplace interventions.

X. CONCLUSION

AI-enabled occupational safety has advanced substantially through computer vision, wearable sensing, IoT, predictive analytics, and multimodal monitoring. However, the reviewed evidence remains concentrated on sensing and algorithmic performance, with considerably less evidence demonstrating reliable workplace operation, effective intervention, or improved worker outcomes. To address this gap, this review introduced a quality- and uncertainty-aware multimodal framework together with a five-level evidence-maturity model linking measurement and algorithmic validity to operational performance, intervention effectiveness, and worker health and safety outcomes.

Effective deployment requires more than accurate hazard detection. AI systems must account for degraded or missing information, communicate uncertainty, support appropriate escalation, and connect recurring risk patterns to the hierarchy of controls. These capabilities should be implemented alongside human oversight, privacy protection, cybersecurity, and responsible data governance. Future research should therefore emphasize external validation and prospective workplace studies that evaluate whether AI-supported interventions translate technical performance into measurable reductions in hazardous exposure, near misses, injuries, and occupational illness.

REFERENCES

[1] G. La Torre, M. V. Manai, S. Meucci, A. Lucente, A. Picerno, S. Ammirati, and S. De Sio, “Artificial intelligence and occupational health and safety: A systematic review,” *Journal of Public Health*, 2026, doi: 10.1007/s10389-026-02738-8.

[2] A. Descatha, D. Hawkins, G. Sembajwe, S. Aloui, E. Fadel, D. Gagliardi, F. S. Violante, M. R. Sim, M. Fadel, and S.-K. Kang, “Artificial intelligence and occupational health: Global umbrella review of applications and limitations,” *Safety and Health at Work*, vol. 17, no. 2, pp. 155–171, Jun. 2026, doi: 10.1016/j.shaw.2026.02.004.

[3] A. M. Vukicevic, M. Petrovic, P. Milosevic, A. Peulic, K. Jovanovic, *et al.*, "A systematic review of computer vision-based personal protective equipment compliance in industry practice: Advancements, challenges and future directions," *Artificial Intelligence Review*, vol. 57, Art. no. 319, 2024, doi: 10.1007/s10462-024-10978-x.

[4] J. Kim, K. Lee, and J. Jeon, "Systematic literature review of wearable devices and data analytics for construction safety and health," *Expert Systems with Applications*, vol. 257, Art. no. 125038, 2024, doi: 10.1016/j.eswa.2024.125038.

[5] M. Wang, J. Chen, and J. Ma, "Monitoring and evaluating the status and behaviour of construction workers using wearable sensing technologies," *Automation in Construction*, vol. 165, Art. no. 105555, 2024, doi: 10.1016/j.autcon.2024.105555.

[6] P. P. Heng, H. M. Yusoff, and R. Hod, "Individual evaluation of fatigue at work to enhance the safety performance in the construction industry: A systematic review," *PLoS ONE*, vol. 19, no. 2, Art. no. e0287892, 2024, doi: 10.1371/journal.pone.0287892.

[7] M. Elrifaee, T. Zayed, E. Ali, and A. H. Ali, "IoT contributions to the safety of construction sites: A comprehensive review of recent advances, limitations, and suggestions for future directions," *Internet of Things*, vol. 31, Art. no. 101387, 2025, doi: 10.1016/j.iot.2024.101387.

[8] A. Jetha, H. Bakhtari, E. Irvin, A. Biswas, M. J. Smith, C. Mustard, V. H. Arrandale, J. T. Dennerlein, and P. M. Smith, "Do occupational health and safety tools that utilize artificial intelligence have a measurable impact on worker injury or illness? Findings from a systematic review," *Systematic Reviews*, vol. 14, Art. no. 146, 2025, doi: 10.1186/s13643-025-02869-1.

[9] B. H. W. Guo, Y. Zou, Y. Fang, Y. M. Goh, and P. X. W. Zou, "Computer vision technologies for safety science and management in construction: A critical review and future research directions," *Safety Science*, vol. 135, Art. no. 105130, 2021, doi: 10.1016/j.ssci.2020.105130.

[10] O. A. Adegboye, T. F. Komolafe, I. M. Joseph, and O. O. Kajero, "Electromagnetic signal classification using deep learning for pipeline structural health monitoring," *International Journal of Scientific Research and Engineering Development*, vol. 7, no. 6, pp. 1317–1322, Nov.–Dec. 2024.

[11] O. A. Adegboye, T. F. Komolafe, and I. M. Joseph, "Physics-aware ML models for pipeline leak detection and quantification," *International Journal of Scientific Research and Engineering Development*, vol. 7, no. 6, pp. 1311–1316, Nov.–Dec. 2024.

[12] O. A. Adegboye, K. M. Udofia, and A. Obot, "Inverted C-shaped slots loaded exponential tapered triple band notched ultra wideband (UWB) antenna," *International Journal of Engineering Research & Technology*, vol. 12, no. 4, Apr. 2023.

[13] J. S. Keränen, J. Ahmad, S. Leggieri, S.-M. Mäkelä, D. G. Caldwell, C. Di Natali, A. Kinnula, and P. Siirtola, "Machine-learning-based fatigue trend analysis on IMU wearable sensor data from construction site workers," *Sensors*, vol. 25, no. 24, Art. no. 7455, 2025, doi: 10.3390/s25247455.

[14] S. Arshad, O. Akinade, S. Bello, and M. Bilal, "Computer vision and IoT research landscape for health and safety management on construction sites," *Journal of Building Engineering*, vol. 76, Art. no. 107049, 2023, doi: 10.1016/j.jobe.2023.107049.

[15] G. A. Vivian, R. A. Bauder, and T. M. Khoshgoftaar, "A comprehensive survey on machine learning for workplace injury analysis: Risk prediction, return to work strategies, and

demographic insights," *Journal of Big Data*, vol. 12, Art. no. 167, 2025, doi: 10.1186/s40537-025-01229-z.

[16] Y. Mao, Z. Ming, X. Yu, and J. Li, "Construction accident narrative classification and risk theme mining using pretrained language models," *Reliability Engineering & System Safety*, 2026, doi: 10.1016/j.ress.2026.112992.

[17] National Institute for Occupational Safety and Health, "Hierarchy of controls," Centers for Disease Control and Prevention, Apr. 10, 2024. [Online]. Available: NIOSH Hierarchy of Controls. [Accessed: Sep. 18, 2026].

[18] International Labour Organization, *Revolutionizing Health and Safety: The Role of AI and Digitalization at Work*. Geneva, Switzerland: International Labour Organization, 2025.

[19] European Agency for Safety and Health at Work, "Worker management through AI: Implications for occupational safety and health," Jan. 20, 2025. [Online]. Available: EU-OSHA publication. [Accessed: Sep. 18, 2026].

[20] European Agency for Safety and Health at Work, "Towards transparency: Smart digital systems for improving workers' safety and health," Jan. 21, 2025. [Online]. Available: EU-OSHA policy brief. [Accessed: Sep. 18, 2026].

[21] S. J. Badhan and R. Samsami, "Artificial intelligence (AI) in construction safety: A systematic literature review," *Buildings*, vol. 15, no. 22, Art. no. 4084, 2025, doi: 10.3390/buildings15224084.

[22] A. A. Fasoyinu, S. Azhar, A. Sattineni, and J. O. Toyin, "Wearable sensing devices for construction safety: Research trends, applications, challenges, and future opportunities," *Automation in Construction*, vol. 179, Art. no. 106424, 2025, doi: 10.1016/j.autcon.2025.106424.

[23] European Agency for Safety and Health at Work, "Artificial intelligence for worker management: Mapping definitions, uses and implications," Jun. 23, 2022. [Online]. Available: EU-OSHA policy brief. [Accessed: Sep. 18, 2026].

[24] European Agency for Safety and Health at Work, "Digitalisation of work," EU-OSHA. [Online]. Available: EU-OSHA Digitalisation of Work. [Accessed: Sep. 18, 2026].

